FDA U.S. Food and Drug Administration

Study Data Specifications and ADaM Review and Comparison

Joy Mele

Office of Biostatistics

CDER/FDA

The views expressed in this talk are those of the author and not necessarily the views of the FDA

Outline

- Study Data Specifications document
  - What is this document?
  - Why did we revise this document?
    - Climate in CDER that suggested changes were needed
    - Process for revising
- Analysis datasets specifications
  - What does the study data specifications document say?
  - How do these specifications fit (or not fit) with ADaM?

U.S. Food and Drug Administration
Protecting and Promoting Public Health

What is the Study Data Specifications document?

"These specifications are for submitting animal and human study datasets in electronic format."

About a year ago …

- Several submissions came into CDER without analysis datasets.
  - Pulmonary drug submission with just SDTM datasets
  - Pain drug submission with only raw data
    - No weighted average, no imputed values, no information on rescue
  - Growth hormone submission with just SDTM datasets

What changed??

**Why were some submissions coming to CDER
with only SDTM-like datasets?**

- We formed an Analysis Dataset
  Submission Working Group
  - Statisticians from every Division of Biometrics

- Goal was to improve communication with
  sponsors to obtain analysis-ready data on
  Day 1 of the review cycle

www.fda.gov?

1. Guidance titled Regulatory Submissions in Electronic Format; New Drug Applications was withdrawn 10/2006

   - Included a section called "General consideration for datasets"

2. Study Data Specifications document was posted 7/2004 and modified 8/2007 to include SDTM links

Description of analysis datasets in 2007 version of Study Data Specifications

<u>**One short paragraph on analysis datasets**</u> that included :

"It is not necessary to provide analysis datasets and programs that will enable the reviewer to directly reproduce reported results using agency hardware and software. Currently, there are no other additional specifications for creating analysis datasets"

And <u>one page on SDTM datasets</u> with links to the CDISC implementation guides for SDTM

How do we quickly get a message out that "yes, we still want analysis datasets"?

- Guidances are usually slow to develop
- Changing our communication avenues via meetings and handouts requires buy-in broadly across CDER
- Webinairs would help but have a limited audience

Our answer was to modify the Study Data Specifications document

Beef up the one paragraph on analysis datasets to at least one page by adding in what we thought was most important from the 1999 guidance

Objectives for Modifying Analysis Datasets Section of Specs

- Our main objective was to bring attention to the need for analysis datasets
  - Not to describe analysis dataset structure in the detail that would be in an implementation guide

- Also wanted to communicate some common needs expressed by CDER statisticians

  - Polled CDER-OB for dataset problems
  - Held a seminar to discuss common problems
  - Wrote the Considerations for Analysis Datasets section with input from OB statisticians supporting all medical divisions

11

Study Data Specification Postings on www.fda.gov

- October 2009 Version 1.5

- January 2010 Version 1.5.1

- Intend to modify quarterly

Study Data Specifications

Analysis Datasets

General Considerations for Analysis Datasets

Dataset Names and Labels

- Specs:
  - Name should be unique within individual studies
  - Internal dataset name = data definition file name
  - Label should clearly describe contents
  - Data definition file should contain at least one dataset labeled as containing <u>primary</u> efficacy data

Where is the primary efficacy data?

D9614C00096 - Data Set Index		
Data Set Name	**Description of Data Set**	**Location**
A120519	Randomization Schedule	~/D9614C00096/crt/datasets/a120519.xpt
AELOG	Adverse Event	~/D9614C00096/crt/datasets/aelog.xpt
A_EVAL	Evaluability	~/D9614C00096/crt/datasets/a_eval.xpt
A_GROWP	Growth Parameter	~/D9614C00096/crt/datasets/a_growp.xpt
CRF_SIGN	Signature	~/D9614C00096/crt/datasets/crf_sign.xpt
CRIT	Eligibility Criteria	~/D9614C00096/crt/datasets/crit.xpt
DEM	Demography	~/D9614C00096/crt/datasets/dem.xpt
DIARY	Electronic IVRS Diary Data	~/D9614C00096/crt/datasets/diary.xpt
DOS	Admin of Investigational Product (Open-label)	~/D9614C00096/crt/datasets/dos.xpt
DOS1	Admin of Investigational Product (Double-blind)	~/D9614C00096/crt/datasets/dos1.xpt
ENDONM	Endoscopy	~/D9614C00096/crt/datasets/endonm.xpt
GIHIS	History of Gastrointestinal Disease	~/D9614C00096/crt/datasets/gihis.xpt
GSO	Physician Global Assessment	~/D9614C00096/crt/datasets/gso.xpt
HISM	Medical History	~/D9614C00096/crt/datasets/hism.xpt
HISS	Surgical History	~/D9614C00096/crt/datasets/hiss.xpt
HIST	Histology	~/D9614C00096/crt/datasets/hist.xpt
HISTOGNM	Histopathological Diagnosis	~/D9614C00096/crt/datasets/histognm.xpt
LAB	Laboratory Assessments SI Units	~/D9614C00096/crt/datasets/lab.xpt
LAB1	Laboratory Assessments - Stool Hemoccult	~/D9614C00096/crt/datasets/lab1.xpt
LABNM	Lab Assessments	~/D9614C00096/crt/datasets/labnm.xpt
LABREF	Laboratory Reference Values	~/D9614C00096/crt/datasets/labref.xpt

Time to Event efficacy data was in DEM dataset

DEM Data Set				
Variable	**Label**	**Type**	**Format**	**Comments**
AGE	Age (Month)	num		
BIRTHDAT	Date of Birth	char		
CENSORD	Censoring Indicator for Time to Discontinuation Due to Any Reason	num	0=Censored	
			1=Event	
CENSORW	Censoring Indicator for Time to Discontinuation Due to Symptom Worsening	num	0=Censored	
TTDISCW	Time to Discontinuation Due to Symptom Worsening (Days)	num		

Questions on dataset label

- Is it implied that a dataset labeled as containing primary efficacy data should contain only primary efficacy data?

NO

- Must all efficacy dataset labels contain the word "EFFICACY"?

NO

- Is it implied that a single dataset must not contain both efficacy and safety data?

NO

Order of the Variables

Specs (p 6)

"The key variables (subject identifier and visit …) should appear first in the datasets. Each subject should be identified by a single, unique subject identifier within an entire application … Subjects enrolled in a primary study and then followed into an extension study should retain their unique identifier from the primary study."

"Core variables should be listed after the key variables and included on each analysis dataset. Core variables include study/protocol, center/site, country, treatment assignment, sex, age, race, analysis population flags (e.g. ITT, safety) and other important baseline demographic variables."

Order conflicts with ADaM? Analysis Data Model Ver 2.1

4.1.3 Ordering of Variables

Ideally, the ordering of the variables in the analysis dataset follows a logical ordering (not simply alphabetic). Refer to the FDA "Study Data Specifications" [7] for more information regarding the ordering of variables in the analysis dataset. It is recommended that the sponsor define a convention for ordering of variables within a dataset and then apply this ordering consistently for all analysis datasets. The ordering of the variables within a dataset should match the order of the variables as presented in the define file.

Visit and Timing Variables

Specs p 6

When a dataset contains multiple records per subject, a variable for relative day of measurement or event and variables for visit should be included. In addition to a protocol-scheduled visit variable, include at least two timing variables; a character variable describing the visit (e.g. WEEK 8) and a corresponding numeric variable (e.g. 8). These two variables are measures of time from randomization.

How about ADaM timing variables?

- Before the new ADaM implementation guide –
  - Visit based on protocol visit numbers, not related to real time on treatment, was commonly used (VISITNUM)

- New ADaM implementation guide
  - AVISIT = **Week 8** & AVISITN = **8**

Date Variables

Specs p7

Dates should be formatted as numeric in the analysis datasets even if dates are in ISO8601 or another character format in the raw data. This formatting will facilitate the calculation of duration.

- Formatting for dates should be the same across all datasets

Reviewer-specified Analysis Datasets?

- Specs (p 5)

  "Prior to submission, sponsors should contact the appropriate center's reviewing division to determine the division's analysis dataset needs."

- Comment

  "ADaM hopes the impression is not given that it is up to the individual reviewer's preference as in the pre-standards era"

24

In the end – the goal is to have analysis-ready datasets routinely submitted to FDA and that may require asking for a dataset with a structure specifically designed for a drug class or type of study

Link to Study Data Specifications

http://www.fda.gov/downloads/Drugs/DevelopmentApproval Process/FormsSubmissionRequirements/ElectronicSub missions/UCM199759.pdf

Path to Document from fda.gov website

Drugs

Information for Industry (under Resources for You)

Electronic Submissions

Electronic Regulatory Submissions

Electronic Common Technical Document (eCTD)